STA_445_Assignment 7

Gabriel Cage-Sepeda

2024-03-28

Load your packages here:

Problem 1:

The `infmort` data set from the package `faraway` gives the infant mortality rate for a variety of countries. The information is relatively out of date, but will be fun to graph. Visualize the data using by creating scatter plots of mortality vs income while faceting using `region` and setting color by `oil` export status. Utilize a $\log_{10}$ transformation for both `mortality` and `income` axes. This can be done either by doing the transformation inside the `aes()` command or by utilizing the `scale_x_log10()` or `scale_y_log10()` layers. The critical difference is if the scales are on the original vs log transformed scale. Experiment with both and see which you prefer.

a. The `rownames()` of the table gives the country names and you should create a new column that contains the country names. *`rownames`

```
infmort <- infmort %>%
  mutate(country=rownames(infmort))
rownames(infmort) <- NULL
infmort
```

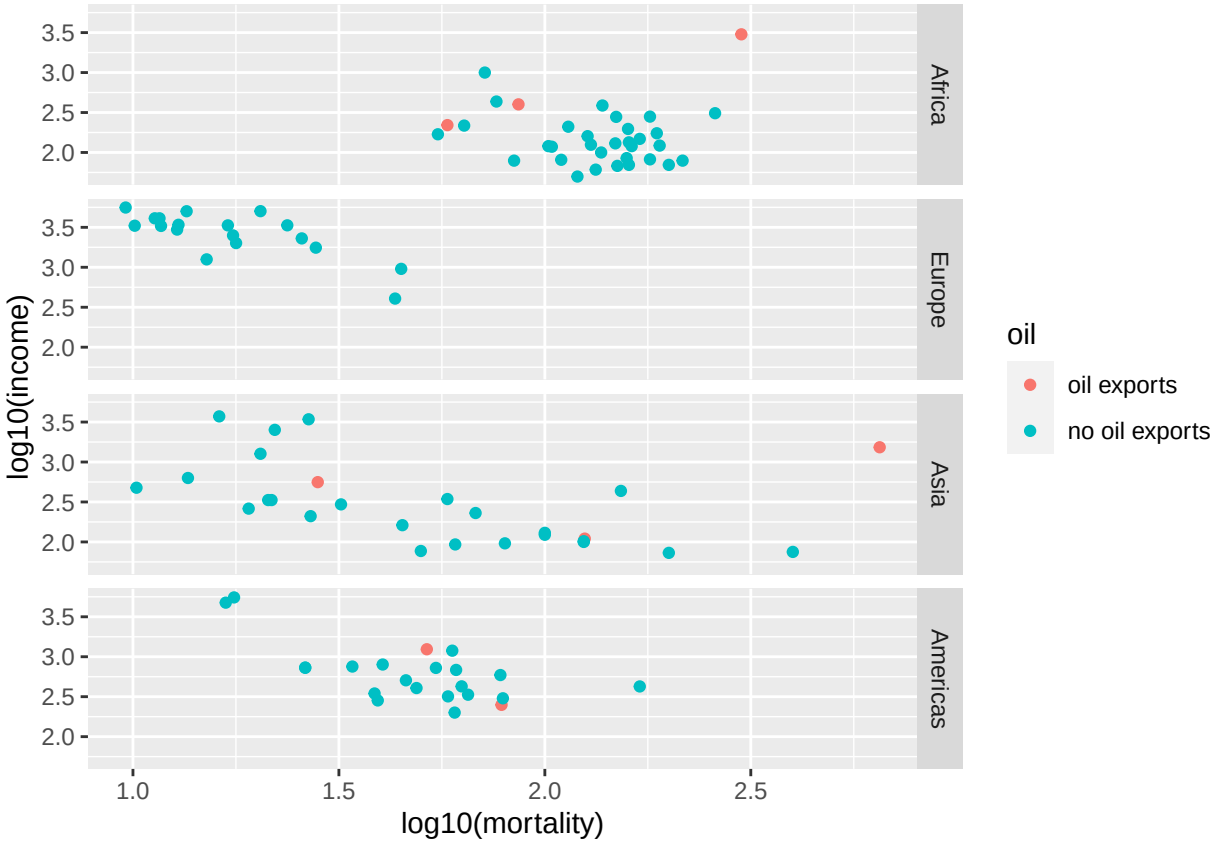
```
##       region income mortality            oil             country
## 1       Asia   3426      26.7 no oil exports Australia
## 2     Europe   3350      23.7 no oil exports Austria
## 3     Europe   3346      17.0 no oil exports Belgium
## 4   Americas   4751      16.8 no oil exports Canada
## 5     Europe   5029      13.5 no oil exports Denmark
## 6     Europe   3312      10.1 no oil exports Finland
## 7     Europe   3403      12.9 no oil exports France
## 8     Europe   5040      20.4 no oil exports West_Germany
## 9     Europe   2009      17.8 no oil exports Ireland
## 10    Europe   2298      25.7 no oil exports Italy
## 11    Europe   3292      11.7 no oil exports Japan
## 12    Europe   4103      11.6 no oil exports Netherlands
## 13      Asia   3723      16.2 no oil exports New_Zealand
## 14    Europe   4102      11.3 no oil exports Norway
## 15    Europe    956      44.8 no oil exports Portugal
## 16    Africa   1000      71.5 no oil exports South_Africa
## 17    Europe   5596       9.6 no oil exports Sweden
## 18    Europe   2963      12.8 no oil exports Switzerland
## 19    Europe   2503      17.5 no oil exports Britain
```

```
## 20  Americas   5523      17.6 no oil exports United_States
## 21    Africa    400      86.3    oil exports Algeria
## 22  Americas    250      78.5    oil exports Ecuador
## 23      Asia    110     125.0    oil exports Indonesia
## 24      Asia   1280        NA    oil exports Iran
## 25      Asia    560      28.1    oil exports Iraq
## 26    Africa   3010     300.0    oil exports Libya
## 27    Africa    220      58.0    oil exports Nigeria
## 28      Asia   1530     650.0    oil exports Saudi_Arabia
## 29  Americas   1240      51.7    oil exports Venezuela
## 30  Americas   1191      59.6 no oil exports Argentina
## 31  Americas    425     170.0 no oil exports Brazil
## 32  Americas    590      78.0 no oil exports Chile
## 33  Americas    426      62.8 no oil exports Colombia
## 34  Americas    725      54.4 no oil exports Costa_Rica
## 35  Americas    406      48.8 no oil exports Dominican_Republic
## 36    Europe   1760      27.8 no oil exports Greece
## 37  Americas    302      79.1 no oil exports Guatemala
## 38      Asia   2526      22.1 no oil exports Israel
## 39  Americas    727      26.2 no oil exports Jamaica
## 40      Asia    631      13.6 no oil exports Lebanon
## 41      Asia    295      32.0 no oil exports Malaysia
## 42  Americas    684      60.9 no oil exports Mexico
## 43  Americas    507      46.0 no oil exports Nicaragua
## 44  Americas    754      34.1 no oil exports Panama
## 45  Americas    335      65.1 no oil exports Peru
## 46      Asia   1268      20.4 no oil exports Singapore
## 47    Europe   1256      15.1 no oil exports Spain
## 48      Asia    261      19.1 no oil exports Taiwan
## 49  Americas    732      26.2 no oil exports Trinidad_and_Tobago
## 50    Africa    434      76.3 no oil exports Tunisia
## 51  Americas    799      40.4 no oil exports Uruguay
## 52    Europe    406      43.3 no oil exports Yugoslavia
## 53    Africa    310     259.0 no oil exports Zambia
## 54  Americas    200      60.4 no oil exports Bolivia
## 55    Africa    100     137.0 no oil exports Cameroon
## 56    Africa    281     180.0 no oil exports Congo
## 57    Africa    210     114.0 no oil exports Egypt
## 58  Americas    319      58.2 no oil exports El_Salvador
## 59    Africa    217      63.7 no oil exports Ghana
## 60  Americas    284      39.3 no oil exports Honduras
## 61    Africa    387     138.0 no oil exports Ivory_Coast
## 62      Asia    334      21.3 no oil exports Jordan
## 63      Asia    344      58.0 no oil exports South_Korea
## 64    Africa    197     159.2 no oil exports Liberia
## 65    Africa    279     149.0 no oil exports Moroco
## 66      Asia    477      10.2 no oil exports Papua_New_Guinea
## 67  Americas    347      38.6 no oil exports Paraguay
## 68      Asia    230      67.9 no oil exports Philippines
## 69      Asia    334      21.7 no oil exports Syria
## 70      Asia    210      27.0 no oil exports Thailand
## 71      Asia    435     153.0 no oil exports Turkey
## 72      Asia    130     100.0 no oil exports South_Vietnam
## 73      Asia     75     400.0 no oil exports Afganistan
```

```
## 74      Asia    100     124.3 no oil exports Bangladesh
## 75      Asia     73     200.0 no oil exports Burma
## 76    Africa     68     150.0 no oil exports Burundi
## 77      Asia    123     100.0 no oil exports Cambodia
## 78    Africa    122     190.0 no oil exports Central_African_Rep
## 79    Africa     70     160.0 no oil exports Chad
## 80    Africa     81     109.6 no oil exports Dahomey
## 81    Africa     79      84.2 no oil exports Ethiopia
## 82    Africa     79     216.0 no oil exports Guinea
## 83  Americas    100        NA no oil exports Haiti
## 84      Asia     93      60.6 no oil exports India
## 85    Africa    169      55.0 no oil exports Kenya
## 86      Asia     71        NA no oil exports Laos
## 87    Africa    120     102.0 no oil exports Madagascar
## 88    Africa    130     148.3 no oil exports Malawi
## 89    Africa     50     120.0 no oil exports Mali
## 90    Africa    174     187.0 no oil exports Mauritania
## 91      Asia     90        NA no oil exports Nepal
## 92    Africa     70     200.0 no oil exports Niger
## 93      Asia    102     124.3 no oil exports Pakistan
## 94    Africa     61     132.9 no oil exports Rwanda
## 95    Africa    148     170.0 no oil exports Sierra_Leone
## 96    Africa     85     158.0 no oil exports Somalia
## 97      Asia    162      45.1 no oil exports Sri_Lanka
## 98    Africa    125     129.4 no oil exports Sudan
## 99    Africa    120     162.5 no oil exports Tanzania
## 100   Africa    160     127.0 no oil exports Togo
## 101   Africa    134     160.0 no oil exports Uganda
## 102   Africa     82     180.0 no oil exports Upper_Volta
## 103     Asia     96      80.0 no oil exports Southern_Yemen
## 104     Asia     77      50.0 no oil exports Yemen
## 105   Africa    118     104.0 no oil exports Zaire
```

b. Create scatter plots with the `log10()` transformation inside the `aes()`command.

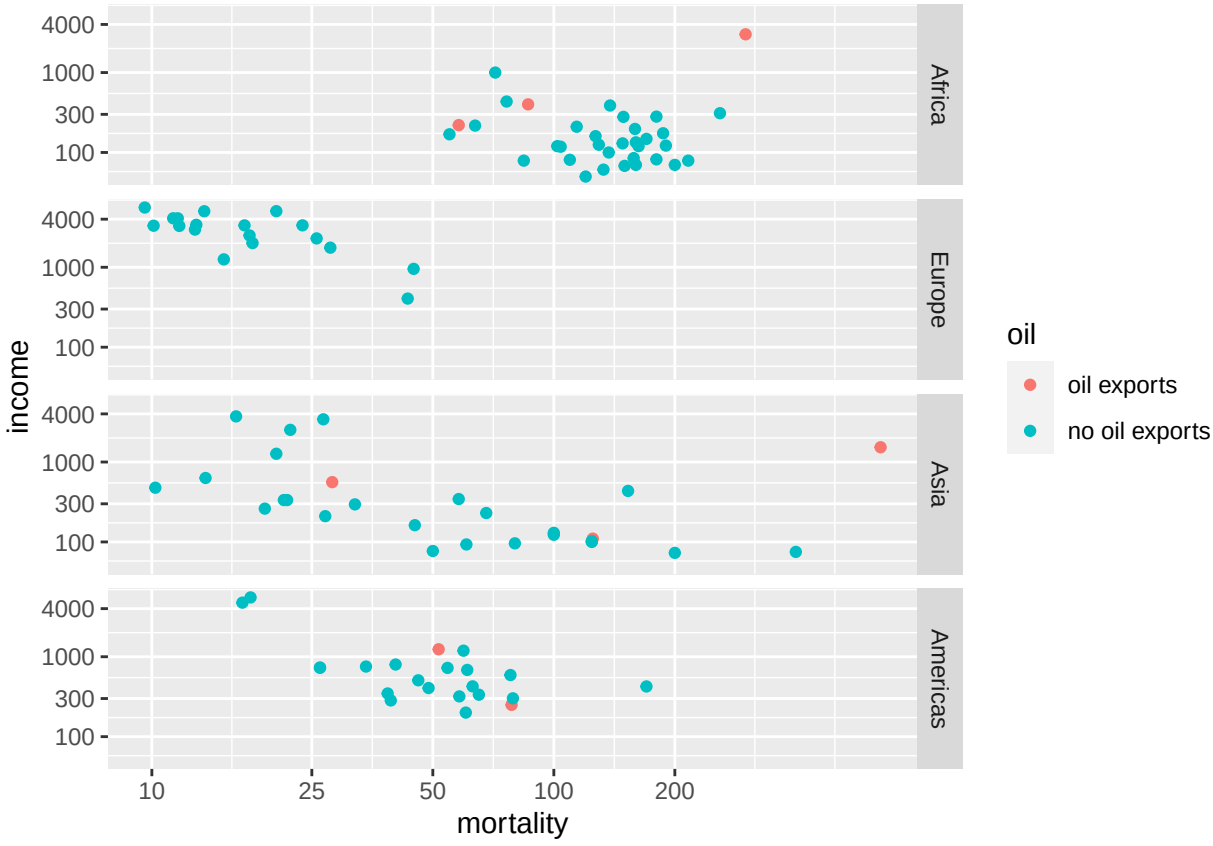
```
ggplot(data=infmort, aes(x=log10(mortality), y=log10(income))) +
  geom_point(aes(color=oil)) +
  facet_grid(region~.)
```



c. Create the scatter plots using the `scale_x_log10()` and `scale_y_log10()`.Set the major and minor breaks to be useful and aesthetically pleasing.Comment on which version you find easier to read.

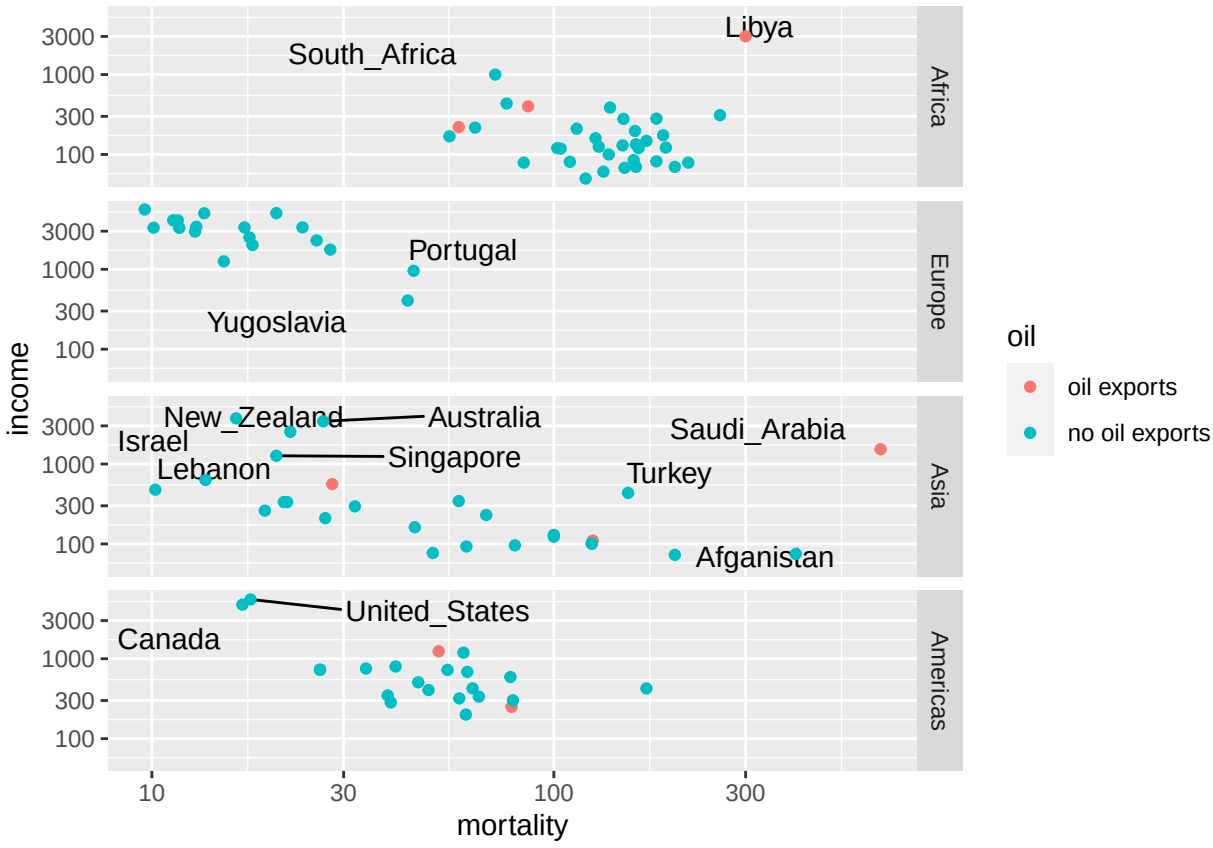
#This version is better because the breaks can be chosen and it uses the original values from the data.

```
ggplot(data=infmort, aes(x=mortality, y=income)) +
  geom_point(aes(color=oil)) +
  scale_x_log10(breaks=c(10, 25, 50, 100, 200)) +
  scale_y_log10(breaks=c(100, 300, 1000, 4000)) +
  facet_grid(region~.)
```



d. The package `ggrepel` contains functions `geom_text_repel()` and `geom_label_repel()` that mimic the basic `geom_text()` and `geom_label()`functions in `ggplot2`, but work to make sure the labels don't overlap. Select 10-15 countries to label and do so using the `geom_text_repel()` function.

```
ggplot(data=infmort, aes(x=mortality, y=income, label=country)) +
  geom_text_repel() +
  scale_x_log10() +
  scale_y_log10() +
  geom_point(aes(color=oil)) +
  facet_grid(region~.)
```



Problem 2

Using the `datasets::trees` data, complete the following:

```
data(trees)
```

a. Create a regression model for $y =$ `Volume` as a function of $x =$ `Height`.

```
trees.model <- lm(data=trees, Volume ~ Height)
trees.model
```

```
##
## Call:
## lm(formula = Volume ~ Height, data = trees)
##
## Coefficients:
## (Intercept)       Height
##     -87.124        1.543
```

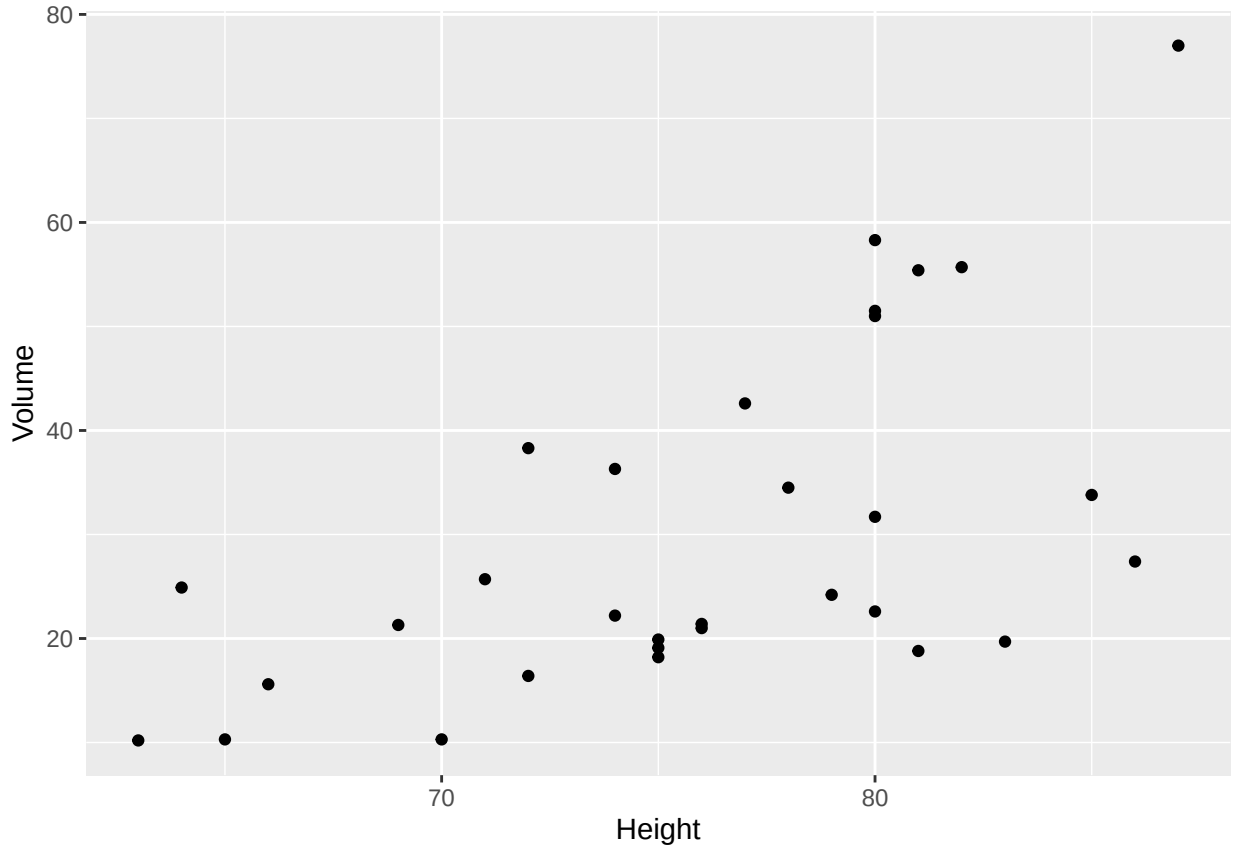
b. Using the str(your model's name) command, to get a list of all the information stored in the linear model object. Use $ to extract the slope and intercept of the regression line (the coefficients).

```
str(trees.model$coefficients)
```

```
##  Named num [1:2] -87.12 1.54
##  - attr(*, "names")= chr [1:2] "(Intercept)" "Height"
```

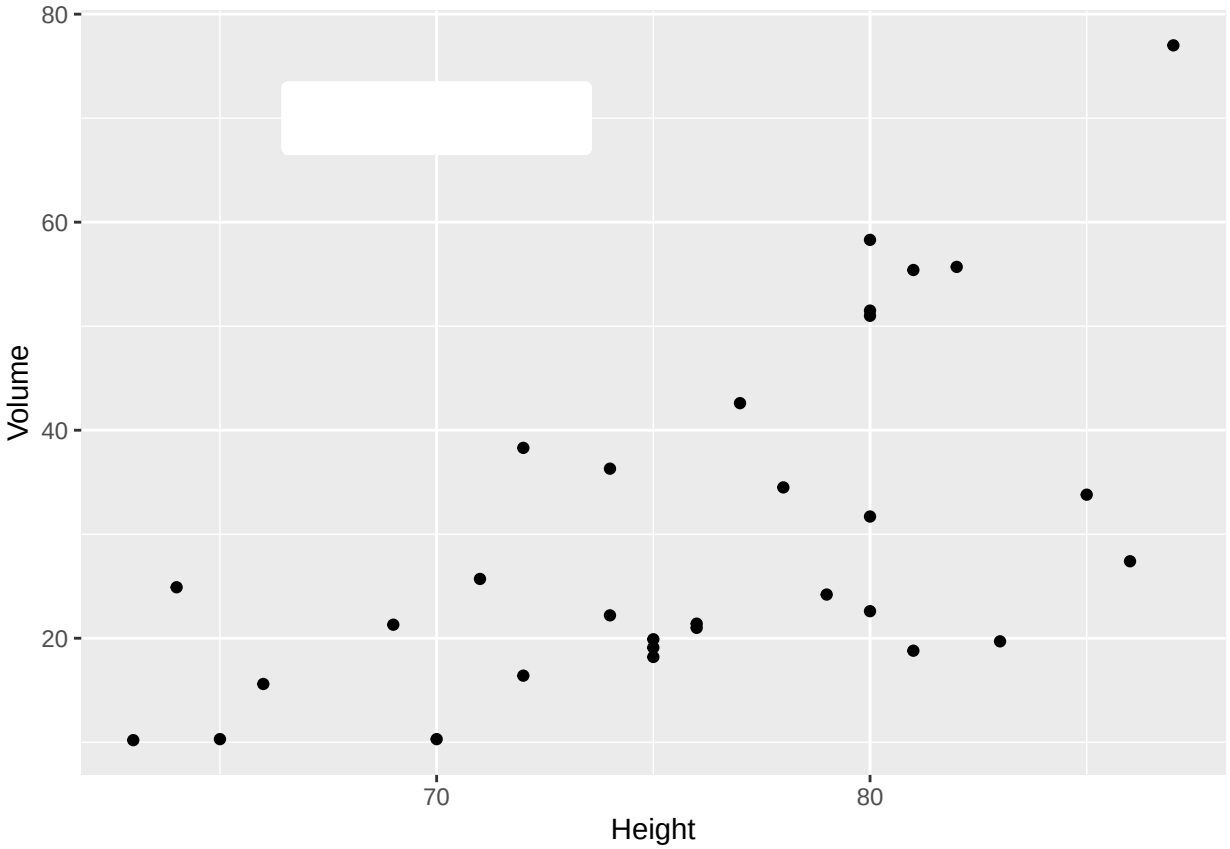
c. Using ggplot2, create a scatter plot of Volume vs Height.

```
ggplot(data=trees, aes(x=Height, y=Volume)) +
  geom_point()
```



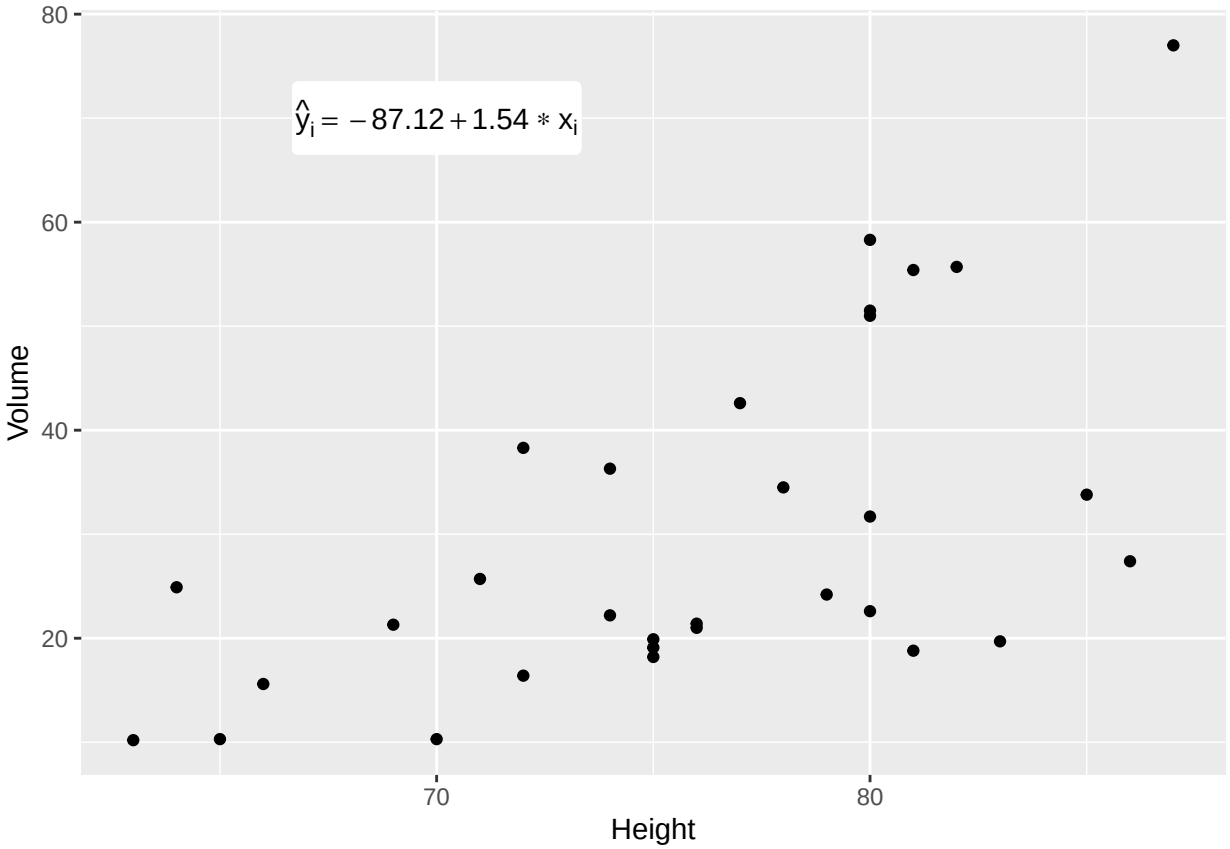
d. Create a nice white filled rectangle to add text information to using by adding the following annotation layer.

```
ggplot(data=trees, aes(x=Height, y=Volume)) +
  geom_point() +
  geom_label(aes(x = 70, y = 70, label = "              "),
             fill = "white", label.size = NA, size = 10)
```



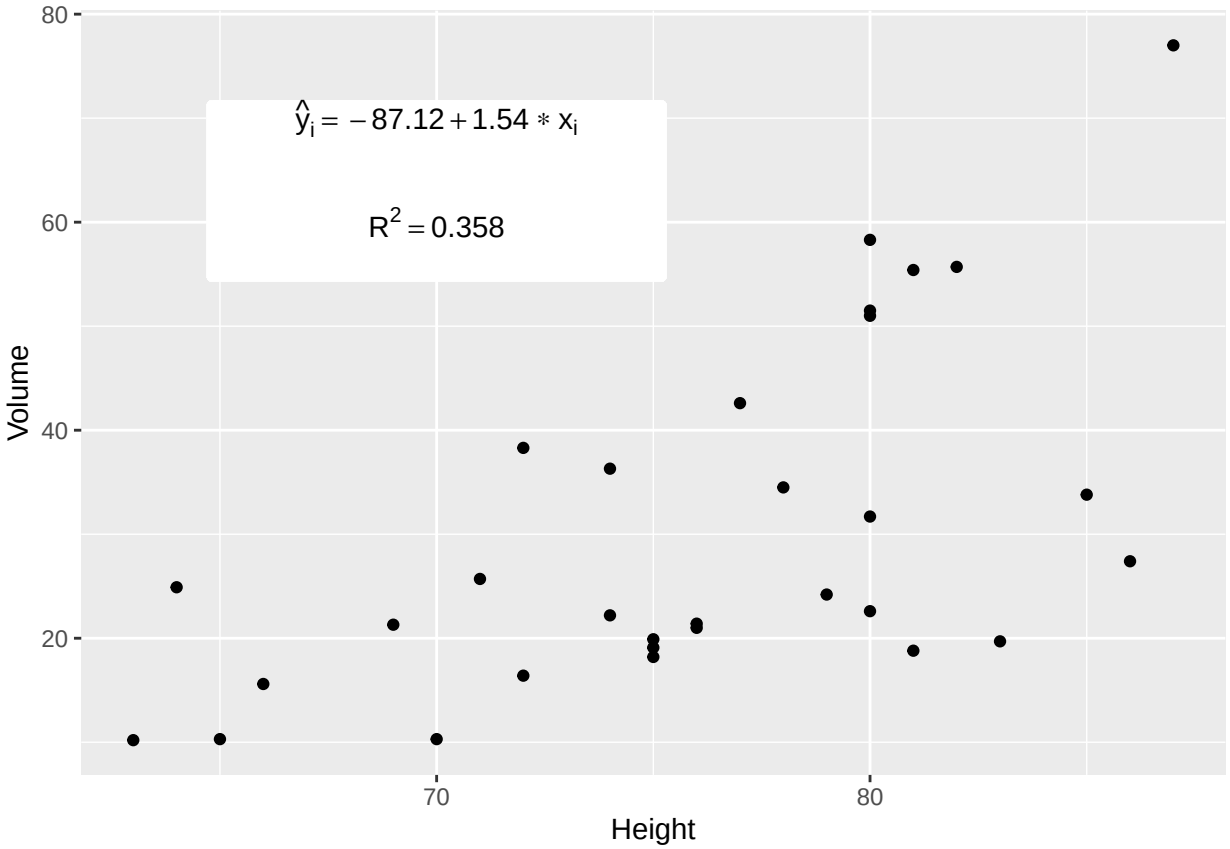
e. Add some annotation text to write the equation of the line $\hat{y}_i = -87.12 + 1.54 * x_i$ in the text area.

```
ggplot(data=trees, aes(x=Height, y=Volume)) +
  geom_point() +
  geom_label(aes(x = 70, y = 70, label = "             "),
             fill = "white", label.size = NA, size = 10) +
  annotate('text', x=70, y=70,
           label=latex2exp::TeX('$\\hat{y}_i = -87.12 + 1.54 * x_i$'))
```



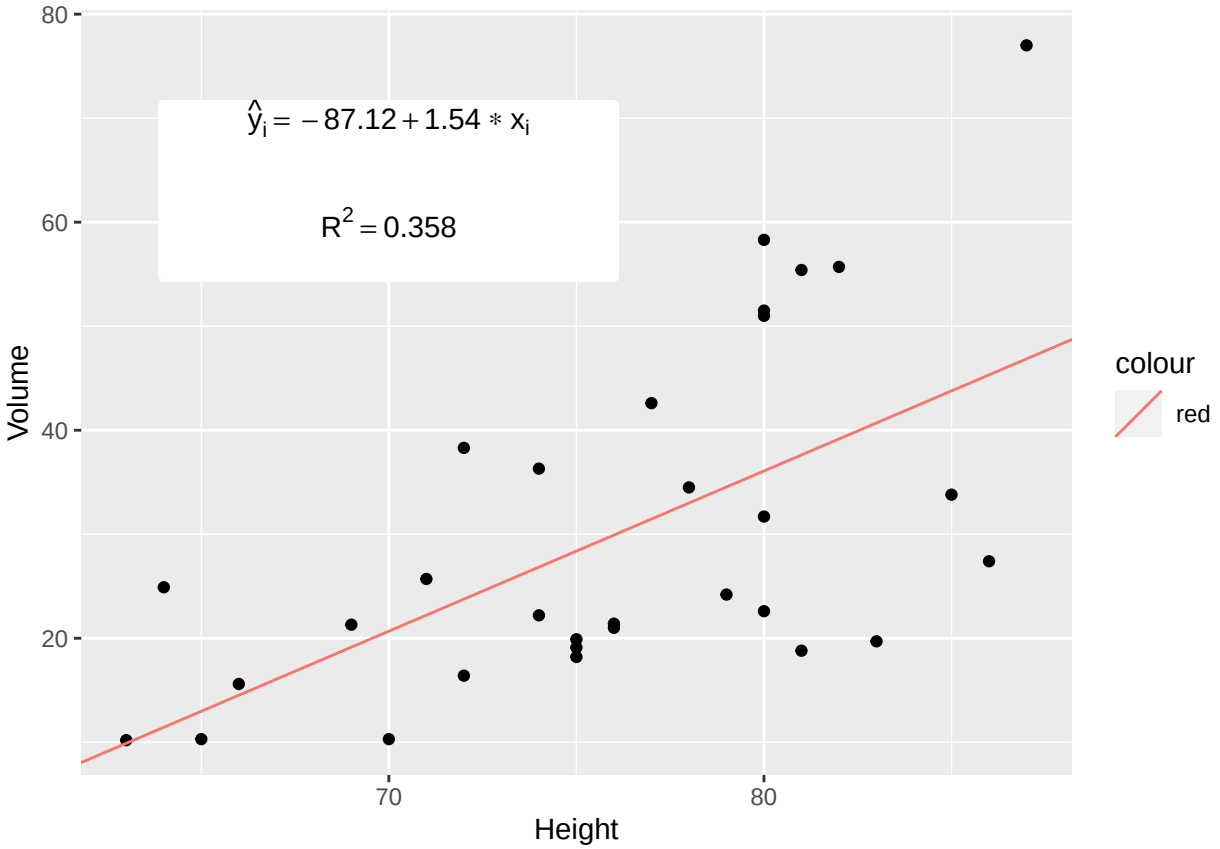
f. Add annotation to add $R^2 = 0.358$

```
ggplot(data=trees, aes(x=Height, y=Volume)) +
  geom_point() +
  geom_label(aes(x = 70, y = 63, label = "       "),
             fill = "white", label.size = NA, size = 30) +
  annotate('text', x=70, y=70,
           label=latex2exp::TeX('$\\hat{y}_i = -87.12 + 1.54 * x_i$')) +
   annotate('text', x=70, y=60,
           label=latex2exp::TeX('$R^2 = 0.358$'))
```



g. Add the regression line in red. The most convenient layer function to use is `geom_abline()`.

```
ggplot(data=trees, aes(x=Height, y=Volume)) +
  geom_point() +
  geom_label(aes(x = 70, y = 63, label = "       "),
             fill = "white", label.size = NA, size = 30) +
  annotate('text', x=70, y=70,
           label=latex2exp::TeX('$\\hat{y}_i = -87.12 + 1.54 * x_i$')) +
   annotate('text', x=70, y=60,
           label=latex2exp::TeX('$R^2 = 0.358$')) +
  geom_abline(aes(slope=1.54, intercept=-87.12, color='red'))
```



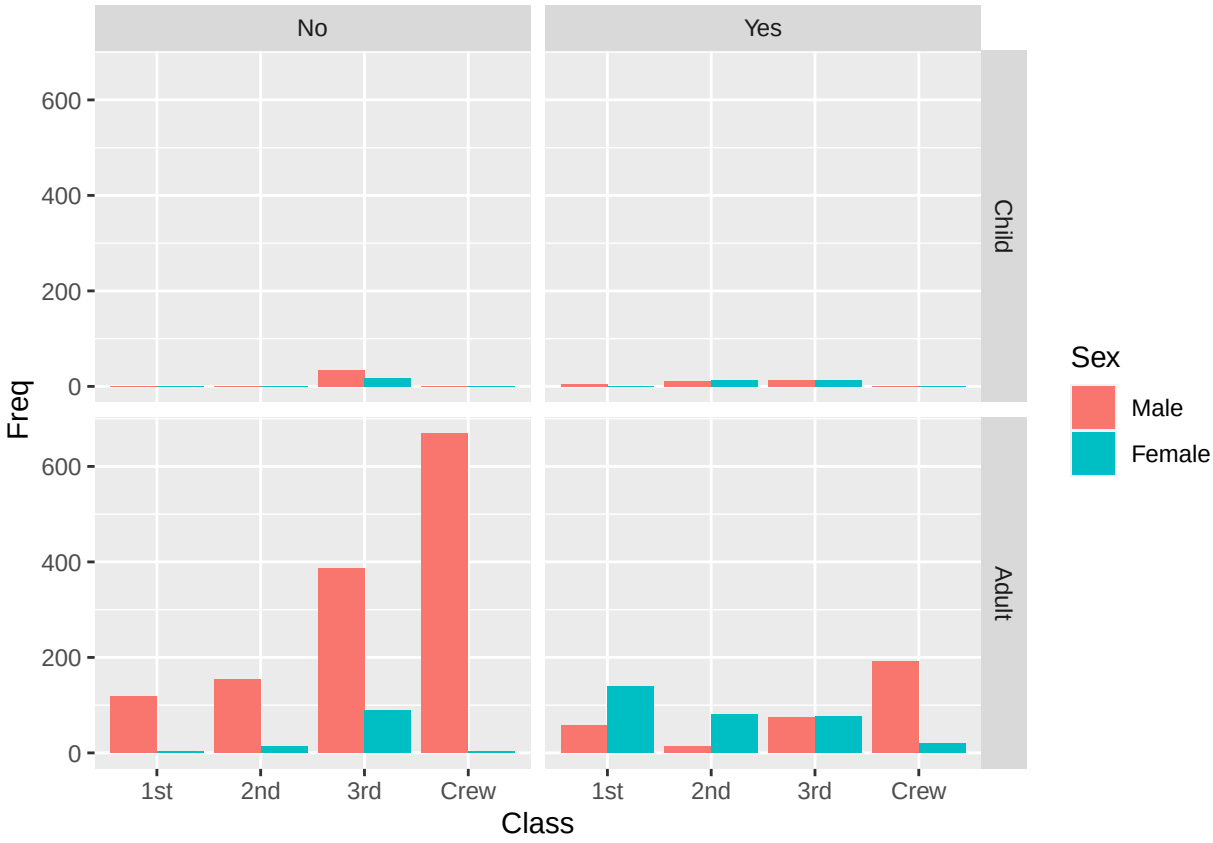
Problem 3

In `datasets::Titanic` table summarizes the survival of passengers aboard the ocean liner *Titanic*. It includes information about passenger class, sex, and age (adult or child). Create a bar graph showing the number of individuals that survived based on the passenger `Class`, `Sex`, and `Age` variable information. You'll need to use faceting and/or color to get all four variables on the same graph. Make sure that differences in survival among different classes of children are perceivable. *Unfortunately, the data is stored as a* `table` *and to expand it to a data frame, the following code can be used.*

```r
Titanic <- Titanic %>% as.data.frame()
```

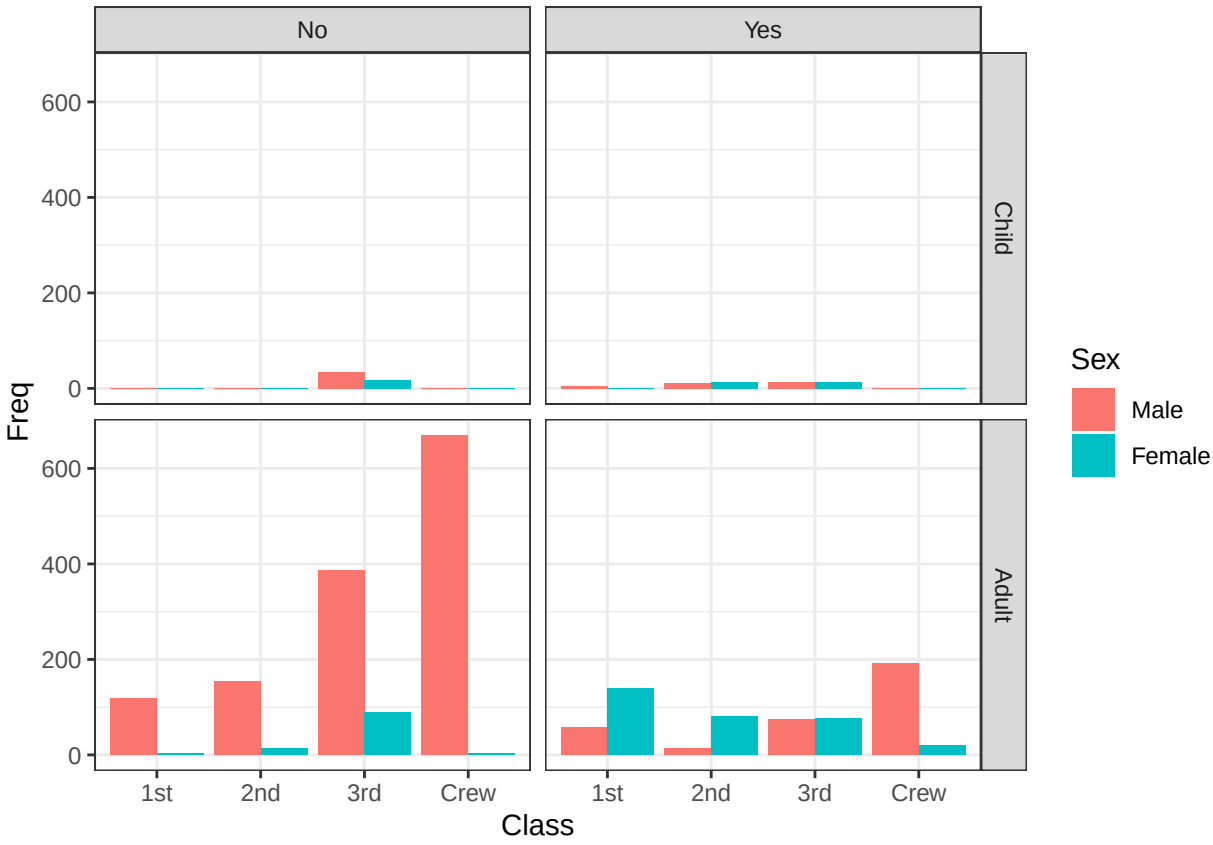
a. Make this graph using the default theme. *If you use color to denote survivorship, modify the color scheme so that a cold color denotes death.*

```
ggplot(data=Titanic, aes(Class, Freq, fill=Sex)) +
  geom_col(position="dodge") +
  facet_grid(Age~Survived)
```



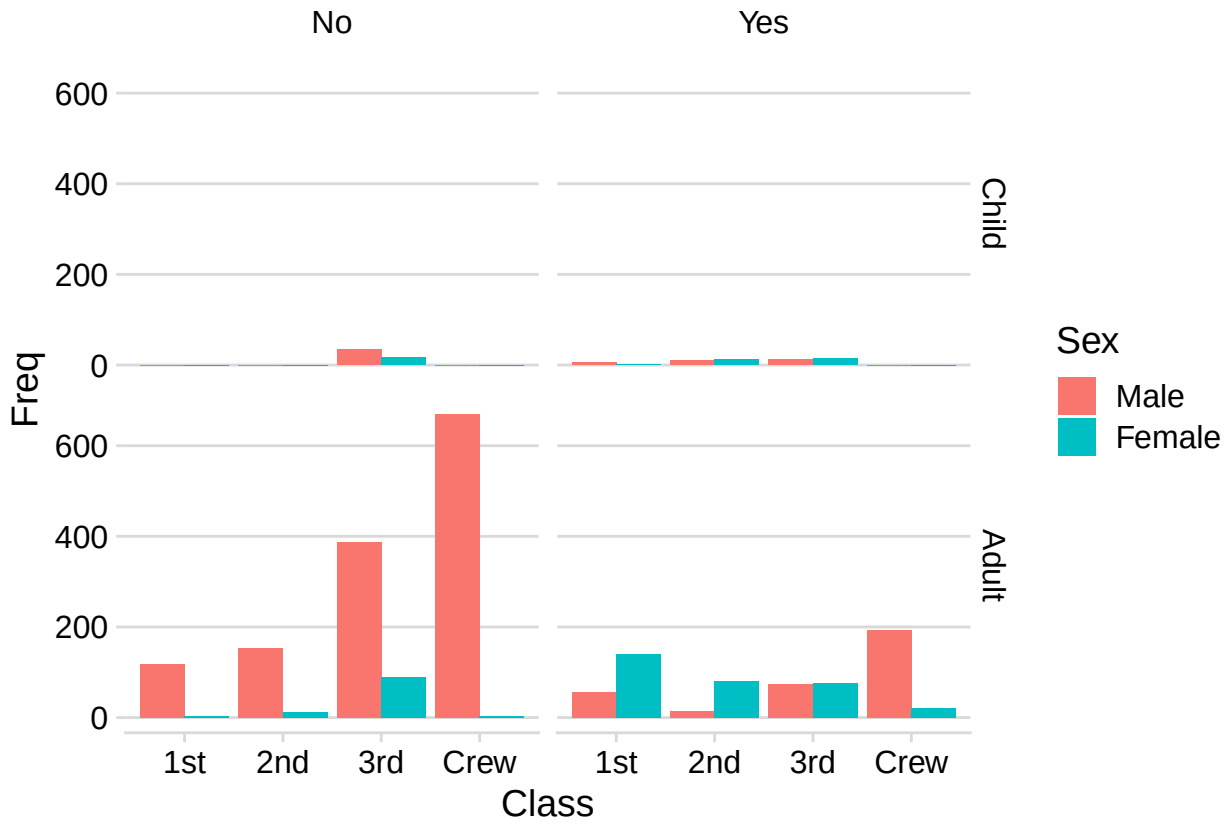
b. Make this graph using the `theme_bw()` theme.

```
ggplot(data=Titanic, aes(Class, Freq, fill=Sex)) +
  geom_col(position="dodge") +
  facet_grid(Age~Survived) +
  theme_bw()
```



c. Make this graph using the `cowplot::theme_minimal_hgrid()` theme.

```
ggplot(data=Titanic, aes(Class, Freq, fill=Sex)) +
  geom_col(position="dodge") +
  facet_grid(Age~Survived)+
    cowplot::theme_minimal_hgrid()
```



d. Why would it be beneficial to drop the vertical grid lines?

The vertical grid lines separate bars that belong to the same variable, easier to read without them.